PRODUCING EDUCATIONAL RESOURCES IN THE "LIBRE WAY": THE EDUKALIBRE PROJECT

Jesús M. González Barahona
Diego Chaparro
Universidad Rey Juan Carlos
Móstoles, Madrid
{jgb,dchaparro}@gsyc.escet.urjc.es

Vania Dimitrova
Chris Tebb
University of Leeds
UK
{vania,chispy}@comp.leeds.ac.uk

Riccardo Mazza
University of Lugano
Switzerland
riccardo.mazza@lu.unisi.ch

ABSTRACT

Recently emerging methodologies for producing educational resources resembling those used in the libre (free, open source) software will radically change the way educational content is developed and used on the web. To fully implement the libre idea, both educational practitioners and students should become actively involved in the creation and distribution of open resources. New architectures are needed to effectively support this process. The paper describes a novel, truly open platform to support the creation of free, collaboratively constructed educational content on the web. The platform has been developed within the Edukalibre project. The Edukalibre system provides easy access to core technologies composed of a control version system combined with conversion tools to produce several convenient formats for each document. Its modular architecture offers many different interfaces to the users. The Edukalibre platform is distributed as libre software.

KEYWORDS

web-based, education, collaborative material, libre software

1. INTRODUCTION

Most educational institutions, especially universities, are exploring the possibilities of applying web-based information systems to education. Currently, such systems are used for re-enforcing or complementing traditional teaching techniques, but, in many cases, new educational paradigms are emerging. New methodologies and architectures are needed to provide effective ways for the creation, sharing, and re-use of educational resources on the web, which can be deployed in a variety of educational paradigms.

This paper discusses recently emerging methodologies for producing educational resources resembling those used in the libre (free, open source) software projects. The libre software development method has tremendously changed the way software is being produced and deployed [12]. Strong communities of practitioners who share experiences, methods, code, and constantly help each other have been built in a number of domains, ranging from operating systems (Debian, FreeBSD, Fedora) to desktop environments (GNOME, KDE), web browsers (Mozilla), web servers (Apache) or office suites (OpenOffice.org) [3].

Similarly, the libre methodology may lead to innovative educational paradigms and will have a great impact on the way the web is used for teaching and learning. Thus, it is extremely important that the educational community becomes aware of the potential of the libre development model. Examining the

connection between libre software development and creation of open content for education is the main objective of the work presented in this paper.

We describe a novel, *truly open* platform to support the creation of free, collaboratively constructed educational content on the web, which has been developed within the Edukalibre project. It is funded by the European Commission under the Socrates/Minerva program, started on October 2003 and is expected to last until December 2005. The project is coordinated by University Rey Juan Carlos (Spain), and includes as partners teams from University of Leeds (United Kingdom), University of Porto (Portugal), University of Karlsruhe, and University of Lugano (Switzerland). Project web site: http://www.edukalibre.org

The paper will first discuss the link between libre software development model and the production of educational materials (Section 2). We will then present, in Section 3, the architecture of the Edukalibre system. Some specific applications of the system will be shown in Section 4. Section 5 will describe the current deployment of the system in university teaching. Finally, in the conclusions, we will sketch some plans for future work.

2. LIBRE SOFTWARE DEVELOPMENT AND EDUCATION

Since early 1980s, the libre software community has shown how software systems can be produced by tightly linked groups of people collaborating over the Internet. Participants in this community have developed over the years a set of practices, procedures and uses for building programs. This set is usually referred as ``the libre software model''. The Edukalibre project explores how this model can be translated to the education camp, and specially to the collaborative development of educational materials. Therefore, we will first outline the key characteristics of this libre software model.

To begin with, it is important to notice that there are nearly as many libre software development models as libre software projects. Variety is one of the most relevant characteristics of the libre software community.

However, there are several specific issues that are found across most libre software projects:

- Frequent and early release of the software, even when it is not fully tested. Eric S. Raymond spells this out as ``Release early, release often''[10].
- Quality by inspection of many individuals, including those outside the group of developers.
- Development is done asynchronously by geographically distributed groups and only using software tools to coordinate[3].
- There is a mixture of voluntary and paid work

The Edukalibre project is exploring how this model can be applied to education, and specifically to the creation and maintenance of educational materials. This means that we explore a model resembling the one of libre software development, but with the particular characteristics of the educational (and more specifically the University) context:

- Educational materials will be located mainly on the Internet.
- Materials will be produced by groups of educators, usually in different institutions, and geographically dispersed.
- Materials will also be used, commented, and maybe modified by students.
- Educators and students will need simple to use, yet powerful, tools to be able to collaborate in the way libre software developers do.
- The public availability of produced materials will enable the collaboration of third parties.
- Several licensing terms will be explored, some of them allowing for publishers to take the materials and distribute them for profit.

Of course, this model is difficult to explore without the proper tools and some user groups ready to test and try it. In the following sections it will be shown how the Edukalibre project is building such tools, and how some test experiences have already started, which are being studied and evaluated within the project.

3. THE EDUKALIBRE SYSTEM

The Edukalibre project has already released the first version of a system that allows for the flexible management (via the web) (using HTTP and WebDAV protocols) of a set of documents. The system is designed to let users choose from a wide set of tools (from easy-to-use word processors, such as OpenOffice.org, to less common XML-based editors), and includes version control facilities and automatic

conversion to many end-user formats (ranging from PDF, ready to be printed on paper, to decorated HTML, suitable for previewing when creating a new version or for reading online).

A diagram of the architecture of the whole system is presented in Figure 1. It shows how the Edukalibre platform is organized around a *document repository*, which is used for supporting collaborative editing, and also provides the basis for a web-based *groupware tool for collaborative authors*. The document repository is built on top of a *version controlled storage system*, with document conversion and management tools written in Python. In addition, some extensions are provided, such as a web based *activity monitoring tool* for teachers. Modules external to the repository can access it directly, using HTTP and WebDAV protocols, or by means of a well defined API. This is the basis of the modular nature of Edukalibre: adding new modules which provide new interfaces to users is a simple process for developers.

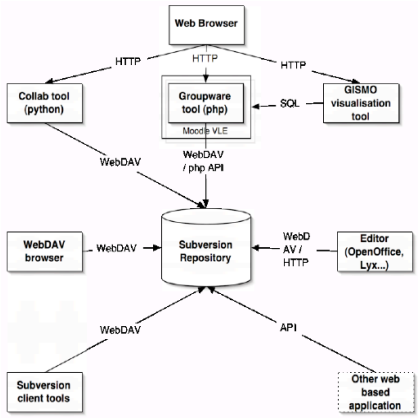

Figure 1 *Arquitecture of the system*

But those components (and others also used) could be easily exchanged for others of similar functionality. In fact, where possible, the project is already providing more than one choice to implement each component, as will be shown later. And of course, the use of standard protocols for the communication and access to the different components of the system has also been a core design requirement.

3.1 Review of Technologies

The system is composed of many different software components, most of them widely used in current libre software projects, ensuring their stability and maturity. When looking at the platform as a whole, we find some of those external modules (libre software code not developed by Edukalibre), some components developed specifically by the project, and some middleware code to link everything together. Figure 2 shows the different components of the collaborative editing subsystem and their main relationships.

The main software components within the Edukalibre project are:

- *Control version system*: Every document created in the system has a main version and a history of older revisions or branches. Each revision can be accessed or modified at any time, and any old revision can be set up as the main version of the document. Currently, the platform uses Subversion as the control version system.
- *Editors and word processors*: The main use of the system is to create documents, therefore a very important component of the system are the editors used to create these documents. Our goal is not to create new editors, but to provide ways in which existing editors and word processors can interact in the most flexible way with the system.

  OpenOffice is an office suite which includes an WYSIWYG word processor. It can edit, in addition to files in its own format, DocBook/XML files (with the help of some code which the project has improved) and Microsoft Word documents. However, any other word processor can be used to modify the documents of the system, if it can manage DocBook/XML or LaTeX, the currently supported document formats. Also a simple web editor is also provided in the system to facilitate the modification of documents using exclusively a web browser.

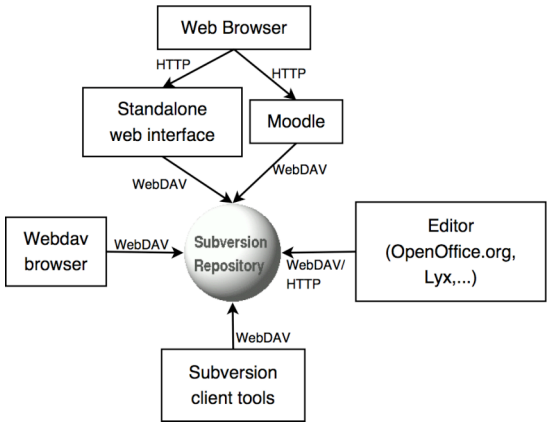

Figure 2 *Components of the collaborative editing tool*

- *Tools for format conversion*: Some other important pieces of software are those in charge of converting documents from the base formats to others, more suitable for reading (such as HTML or

PDF). Most of the conversions from XML formats are performed using XSLT stylesheets. The XSLT processors used are xsltproc and xalan. Conversions to PDF and Postscript are done using TeX/LaTeX and Ghostscript.

- *Web interfaces*: One of the goals of the project is to allow for as many different interfaces to the system as possible, and to let users decide which are their preferred writing tools. In addition to accessing the system directly with a text editor plus a Subversion client, other interfaces are already provided like a collaborative editing system (COLLAB), which is a simple Python-written web interface. The same collaborative editing system has also been integrated as a component of the libre software Moodle LMS, providing all the additional functionality of a learning environment to support the collaborative editing process. There is also a PHP groupware application written as a component of Moodle, called ConDOR which aim is to examine the effectiveness of the tools within a simple groupware application.
- *Communication protocols*: Standard protocols are used to connect the different components. They are the ubiquitous HTTP [1], and WebDAV [4] (an extension to HTTP tailored to the needs of collaborative editing and management of files on remote web servers).

The Edukalibre system offers a PHP API to allow the easy integration of any PHP application with it. It uses the libre software VersionControlSVN package, which is a part of the PEAR PHP libraries. The Edukalibre PHP API builds on this to provide simple XHTML form generation for interacting with the Python scripts that control the repository and a set of file management tools for folder creation, file deletion, history extraction and functions that return the differences between certain versions of a given document.

3.2 Formats

The central elements of the Edukalibre system are the documents. Every document can be accessed in many different formats, but it is convenient to distinguish between two kinds of those formats:

- Base formats that can be the inputs for the whole system. A document can be produced or modified in any of these base formats. At the moment the base formats are: DocBook/XML, OpenOffice.org and LaTeX.
- End-user formats, which are automatically generated by the system from the base formats. These include PDF or Postscript (both suitable for printing), XHTML in several flavors or plain text.

It is important to notice that only base formats of the documents are stored in the control version repository. The end-user formats are stored independently, in the end-user repository, since they can be generated from the base formats. The schema for conversions between formats is given in Figure 3.

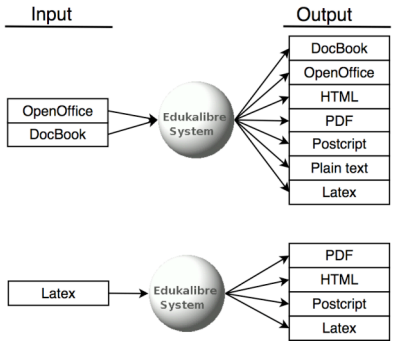

Figure 3 *Document formats*

3.3 Repository Manager

The repository manager is the core of the Edukalibre platform, and also the heart of the collaborative editing tool (see section 4.1). It provides the main functionality of the system, being in charge of storing the documents, creating the end-user formats, and providing information about the documents stored in the system. It can be split in two parts: the *core repository*, and the *document conversion and metadata extraction system*.

- The core repository is basically a storage system in which documents, together with information about them, are stored. The document repository is a subversion repository, with additional plugins to perform some core tasks when a document is uploaded. The second element is just an HTTP server, which stores the end-user formats of all versions for all documents.
- The document conversion and metadata extraction system provides the functionality of the repository manager. The system is launched when a new document or a new version of an existing document arrives to the repository manager via any of the different interfaces shown in Figure 1. From there on, it validates the file, stores the file in the subversion repository, and converts the file to the end-user formats, which are later stored in the end-user repository. At the end, some

information is extracted from the document, and also stored in the system (such as title or authors).

This system also provides methods used by the different interfaces to extract metadata information about the content of the repository manager: listing of documents stored in the system, information about history of each document (versions, authors, dates, etc), logs about the conversion of each end-format for each document and general properties about each document (base format, title, abstract, etc).

4. APPLICATIONS

In the following subsections, some applications that are part of the Edukalibre system are described. All of them show both how the functionality of the system can be integrated with external tools, and also how the architecture is modular enough to make room for may different ways of interaction.

4.1 COLLAB: a Python-driven Web Interface

This is one of the possible interfaces to the system, and allows the users to perform some of the common tasks using just a web browser. A screenshot of it can be seen in Figure 4.

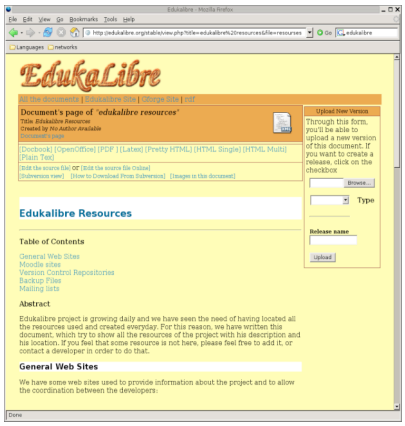

Figure 4 *Collaborative editing tool interface*

The main page of the interface consists in a list of documents created in the system, with some information about each of them. The interface provides a way to show complete information about each document, and to create new documents. An RDF channel with information about the documents recently modified or created is also provided. The information for each document includes:

- Listing of its version history, with the release date and the comments from the release author.
- Listing of links to all end-user formats for every version.
- Editing and downloading options for each version, including the possibility to download the base format, modify it and upload the new version or to edit it on-line.
- Forms to update the document, uploading a new version (written with a standalone editor such as OpenOffice.org).
- Decorated HTML version of the latest version of the document as a quick preview.

The interface also provides a form to upload a new document to the system.

4.2 ConDOR: An Intuitive PHP-driven Groupware Application

One of the core deliverables of the project was to create a *user friendly* working environment which would allow us to evaluate the effectiveness of web-based version controlled collaborative authoring.

To this end, a bespoke groupware application was produced, which ran as a modular component of the libre software learning management environment Moodle, see Figure 5. The groupware was written in PHP, and made use of the PHP subversion API which we created to allow easy access to the document repository for PHP web application developers.

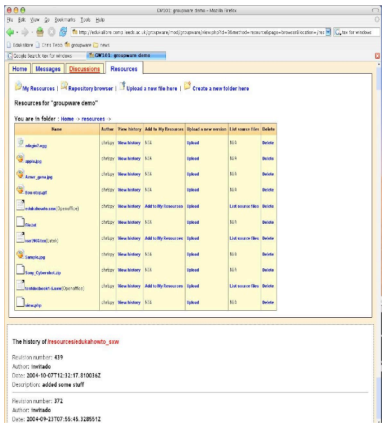

Figure 5 *Groupware tool for constructing of dynamic open resources (ConDOR)*

The requirements specification for the groupware was taken from a questionnaire that was filled in by all partners of the project, enquiring into the scenarios in which the platform would be tested.

The majority of the official evaluation scenarios were based around a teacher assigning collaborative group coursework for students, and the students using the groupware tool to produce shared resources, take part in discussions about

the work and produce a single final document for submission. The basic requirements of the users were:

- Access to shared resources
- Access to discussion facilities
- Easy version control

The groupware application allows easy ``explorer style'' access to the document repository, allowing intuitive navigation of the file and folder structures, and simple tools for uploading files and creating new folders.

With the current prototype, each groupware is linked to a fixed repository, but the system will allow the users to browse the content in multiple arbitrary repositories. To prevent excessive repeated navigation of large folder structures, a ``My Resources'' section is provided within the groupware resource area to allow authors to bookmark documents deep within the resource tree that they have a current active interest in.

From either the main resources browser or the ``My Resources'' bookmark area, authors can download a document in any supported format (currently XHTML, PDF and plain text), upload a new version of the document, browse the history of the document and add/remove documents from their ``My Resources'' list.

The repository will allow authors to add any type of file, and put it under version control. Extra features such as format conversion are only available for OpenOffice and LaTeX documents, but trial users have been using the repository to manage many kinds of file.

It is the aim of the developers to provide as unrestrictive a set of authoring tools as possible. Currently, the system supports offline authoring of any file type by allowing users to download a file, work on it using a local application and upload a new version. The next phase of development will see online editing and simple visual change management (similar to the programmers tool diff) of XHTML, LaTeX, OpenOffice, DocBook/XML and Microsoft Word documents.

4.3 GISMO: a Graphical Interactive Students MOnitoring tools

As shown in Figure 2, the Edukalibre system is currently linked to the Moodle Learning Management System. Educational materials collaboratively produced by using COLLAB or ConDOR are made available to the students through Moodle.

Student tracking data provided by the Learning Management System is a valuable source of data that can be used by the instructor not only to check students activities, but also to improve quality of the materials. For instance, an instructor may check which part of the course's materials are most and less accessed by the student, and then perform further investigations to understand whether the students found these parts difficult to understand or not.

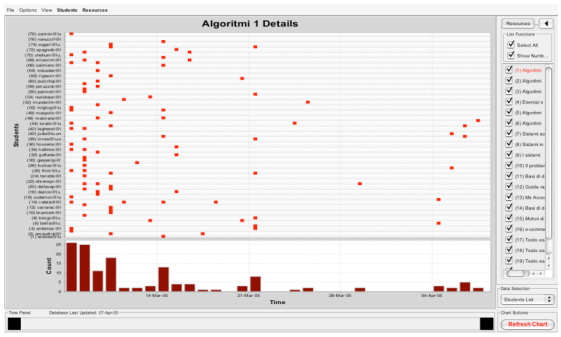

Figure 6 *Representation of the students accesses to a particular content material of the course*

Student tracking data is complex and is usually organized in some form of a tabular format, which is in most of the cases difficult to follow and inappropriate for the instructors' needs [8]. For this reason, a graphical and interactive tracking tool, GISMO, was implemented as an application that interfaces with the Edukalibre system. GISMO uses the students' tracking data as source data, and generates graphical representations that can be explored and manipulated by course instructors to examine social, cognitive, and behavioural aspects of distance students. From a technical point of view, GISMO is application that runs in conjunction with Moodle, and it is delivered through the Web using a Java Applet. However, it can be adapted to support other learning platforms. In fact, a software API is committed to fetch some data that is usually present in a wide range of platforms like Moodle, Claroline, Fle3, Mimerdesk, etc., such as discussions, accesses to the course, marks students receive in quizzes, and so on.

Instructors can investigate the users' reading of course materials, which can give invaluable feedback on the quality of collaboratively constructed resources and their practical use by students. Figure 6 shows an example of GISMO visualization that can be used by instructors to get insights of how popular is the content material to the students of the course. It shows, for a specific part of the course content, in which days the students accessed this material of the course and how many times. The chart on the top illustrates accesses for each student. The histogram on the bottom represents the sum of accesses made by students to the

resource on each specific date. Several other visualisation have been implemented, there are thoroughly described in [9].

5. DEPLOYMENT OF THE EDUKALIBRE SYSTEM IN UNIVERSITY TEACHING

The Edukalibre project is scheduled in two main phases. The first one is devoted to the building of a proof-of-concept system which can be used to explore the new development models of educational materials. At the moment of writing of this paper, the project is in the transition from the first to the second phase. A deployable version of the system described in the previous sections has already been completed and is being deployed in real settings.

We are now conducting formative evaluation [7] to examine the system's functionality and usability. It includes both *expert inspection* and *user trails* of the three applications: COLLAB, ConDOR, and GISMO.

The *expert inspection* is based on a framework for usability evaluation of e-learning web applications, called MILE [11], developed by a team of Human-Computer Interaction researchers from the Politechnico di Milano and the University of Lugano. The inspection is driven by typical educational scenarios, which were defined based on a questionnaire involving university teachers from all Edukalibre partner institutions.
Twelve experts, including teachers and human-computer interaction specialists, have been asked to inspect the Edukalibre applications, and to fill in MILE usability sheets. This evaluation is on-going.

Formative evaluation can be addressed properly only if inspection is combined with real user trials. In line with the Edukalibre objectives, the initial user trials consider the creation of open educational content by both user groups - teachers and students. During the last two months the Edukalibre applications are being used within partner institutions in the following situations: several authors working together to prepare educational content, researchers creating a repository of papers and collaboratively written reports, and students creating repositories on a common topic of their dissertation projects.

Some preliminary comments from inspectors and users have already highlighted issues that are being fixed by the Edukalibre development team, such as:

- *flexibility* - which refers to the robustness of the system to deal with a variety of file formats, at the moment, version control is being fully enabled for OpenOffice and Latex documents;
- *effectiveness* - which refers to the applicability of the system in learning scenarios, and required several iterations to ensure that applications complied with the tasks they were intended for, i.e. collaborative writing by teams of teachers or students (COLLAB and ConDOR), resource management for group projects (ConDOR), and monitoring of student activities (GISMO);
- *efficiency* - which refers to how quickly can the users perform their tasks and resulted in fixing some cumbersome operations and long navigation menus;
- *user satisfaction* - which refers to creating subjectively pleasing systems, and required the fixing of some design and aesthetic issues.

During the second Edukalibre phase, a summative evaluation of the improved system will be conducted in real settings at partner institutions. These include next situations using COLLAB, ConDOR and GISMO:

- Students working in groups to take a decision on an environmental case and write a group report to justify their decision - this is being prepared by the partners from Karlsruhe;
- Students researching and producing content on a new topic, including collaboratively written review and a collection of resources on the web - this is being prepared as part of a module on Personalization and user-adaptive systems at Leeds;
- Educationalists from geographically remote locations constructing collaboratively educational content on statistical methods and operational research, which includes teachers and researchers from Madrid, Prague, and Porto.

6. CONCLUSIONS AND FURTHER WORK

The Edukalibre project is exploring the field of collaborative production of educational materials from a novel point of view: to which extent the common practices of the libre software community can be translated to the educational content domain. For this exploration, the project has developed (mainly by reusing already

existing components) a web-based system that aims to provide some of the functionality available in version control systems, but specifically targeted to the producers of educational documents. With this target population in mind, an effort has been done to minimize the learning effort for using the tools, yet maintain as much functionality and flexibility as possible.

There are other projects which deal with some of the aspects important for Edukalibre. Among them, it is worth mentioning Apache Forrest, which is a documentation framework which can be used to render source content in several formats (including a subset of DocBook/XML) into a complete, customized website. Some ideas of the architecture for converting and managing formats are taken from it. It is also relevant to consider BSCW [2], a cooperative application which supports cooperation through small repositories in which users can upload documents, hold discussions and coordinate their work with other user. However, no version control in the sense provided by Edukalibre, or complex conversions to several formats is provided. In some sense, BSCW could be the basis for some of the tools that Edukalibre misses to replicate some of the functionality provided by GForge (such as discussions or coordination tools). Another system with some resemblances is Connexions [5], which offers a collection of free educational materials and a set of free software tools to help authors, instructors and learners (but is more a collection than a supporting system, like Edukalibre). Another example is the SERUM [6] system, which proposes a model of customized repositories for collaborative authoring of web content, and in this is similar to the Edukalibre system presented here.

Currently, the Edukalibre system is entering its second year, when in addition to some evaluation activities, new lines of work are to be open. Among them, some of the most relevant are the exploration of more complete version control patterns (including the use of private branches, for instance for groups of students maintaining their own customized and commented version of a document); the improvement of the different interfaces and the integration with OpenOffice.org; the improvement of the integration with LMS systems such as Moodle; and the integration with wiki-like tools.

In the future, more work can be done in the direction of integrating more and more subsystems common in libre software development, such as bug tracking systems (for unresolved problems with document content), activity trackers, release management systems, etc. In the end, it would be interesting to consider the construction of hosting sites for the creation and management of documents and other types of materials, in the same sense that sites like SourceForge provide support and hosting for libre software projects.

REFERENCES

[1] R. Fielding et al, 1999, Hypertext transfer protocol - HTTP/1.1, RFC 2616. *Technical report, Network Working Group, IETF.*

[2] BSCW: Basic Support for Cooperative Work, http://bscw.fit.fraunhofer.de/.

[3] Daniel M. German, 2003, The GNOME project: a case study of open source, global software development, *Software Process Improvement and Practice*, pages 201-215

[4] Y. Goland et al, 1999, HTTP extensions for distributed authoring - WEBDAV, RFC 2518. *Technical report, Network Working Group, IETF.*

[5] Connextions: Sharing knowledge and building communities. http://cnx.rice.edu/.

[6] J. Kovse et al, 2001, Supporting collaborative authoring of web content by a customizable resource repository. *Proceedings GI/OCG-Jahrestagung Informatik 2001, Workshop "Web Databases" (Wien)*, pages 358-367

[7] M.A. Mark and J. Greer, 1993. Evaluation methodologies for intelligent tutoring system. *Journal of Artificial Intelligence in Education*, Vol. 2/3, No. 1, pages 129-154

[8] R. Mazza and V. Dimitrova, 2004, Visualising student tracking data to support instructors in web-based distance education. *Proceedings of the 13th International World Wide Web Conference (WWW 2004) - Alternate Educational Track*, pages 154-161.

[9] R. Mazza, C. Milani. GISMO: a Graphical Interactive Student Monitoring Tool for Course Management Systems. *T.E.L.'04 Technology Enhanced Learning '04 International Conference*. Milan, 18-19 November 2004.

[10] Eric S. Raymond, 1997. The cathedral and the bazar. *First Monday,* http://www.firstmonday.dk/issues/issue3_3/raymond/

[11] Luca Triacca et al, 2004, Mile: Systematic usability evaluation for e-learning web application. *Proceedings of ED-MEDIA04 World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Lugano, Switzerland.

[12] Ahmed E. Hassan et al, 2001, Software engineering research in the bazaar. *Proceedings of the 2nd Workshop on Open Source Software Engineering at the 24th International Conference on Software Engineering.*